



SESSION 1

PATTERNS & PREDICTIONS

Matthew Caldwell, COMPASS, UCL

observing patterns, coming up with mental models of what gives rise to those patterns, then making predictions based on that and maybe acting on those predictions, is central to a lot of what people do everyone does it, all the time



doctors look at patterns of symptoms, come up with diagnoses and treatment plans
town planners look at traffic patterns and try to predict congestion or decide where to build a bypass
businesses observe patterns of trade and flows of capital and predict market opportunities
astrologers see patterns in the stars and predict long journeys and tall dark handsome strangers
the patterns may be spurious, the predictions may be wrong, but we can't help doing it

obviously this isn't just restricted to legitimate actors, it applies to criminals and terrorists too



you notice old Doris is out every Thursday at Bingo, predict her house will be empty that night next week, choose to burgle it then



maybe you observe that jewellers do a roaring trade in February, it being Valentine's Day and all



and predict that come the end of the month the security van will be collecting huge wads of cash, giving two weeks...



to get the lads out of retirement on the Costa del Crime for one last heist...

**Like computers or the internal combustion engine,
AI is a general-purpose technology that can
be used to automate a great many tasks,
including ones that should not be undertaken
in the first place.**

—Agüera y Arcas, Todorov & Mitchell, 2017

many of the benefits of AI come from ability to do such pattern-finding and exploitation better, faster and on a larger scale

of course, that also applies to the crimes



much early AI focused on things people find difficult, like chess, or proving mathematical theorems
things intelligent people do, so doing them is by definition intelligent
also, conveniently, things that tend to have well-defined rules that can be explicitly expressed
so: figure out how people do it, write a program that replicates that, et voila: an intelligent machine

but it turns out that much of the interesting stuff in AI
is stuff that humans *don't* find difficult
-- but that we have absolutely no idea how we do



eg talking, walking, reading, recognising faces, recognising danger, jumping to conclusions, arguing, joking, persuasion, lying

because we don't know how we do this stuff, it's really difficult to codify
often even to define what it is
let alone write a program to do it

and also, because it comes naturally, we usually have no idea how hard it is
which is usually **very**

so how can we tell a machine to do something when we don't know how it's done?

well, a lot of the way humans learn things is by example



we don't learn what cats are by reading a detailed specification document...



...we do it by meeting cats.

so one key strand of AI involves just that: looking at examples -- usually *lots* of examples -- and finding patterns in them
this is, very broadly, what we call...

Machine Learning

"Machine Learning", an AI paradigm that is currently quite dominant

Machine Learning

ML isn't the whole of AI, but...

Deep Learning

Machine Learning

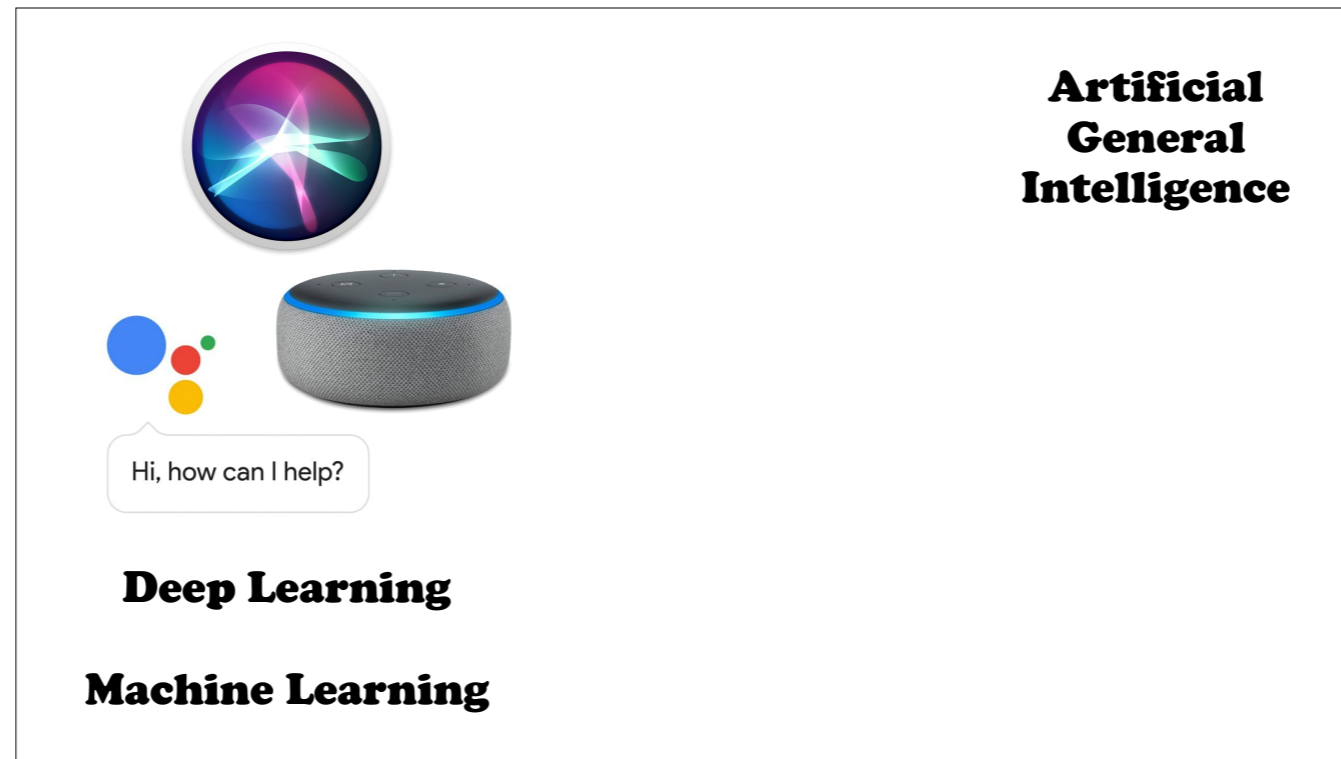
along with Deep Learning — which is basically the same thing with more complex underlying models
— it is responsible for most of the current interest in it
and it provides the technical solutions that underpin



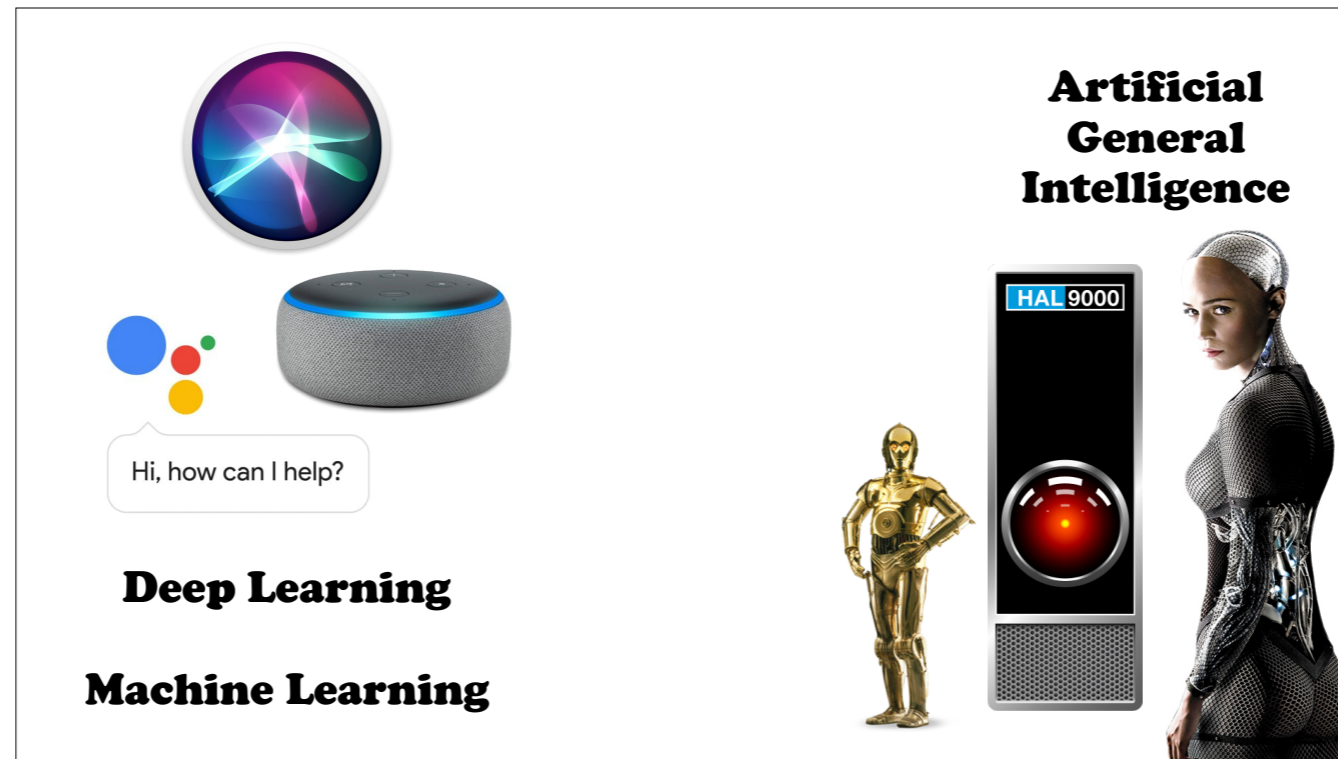
mass market products like Siri and Alexa.

ML & DL solve difficult problems, but they are discrete, specific, well-defined problems,

it is important to distinguish this from what is often called

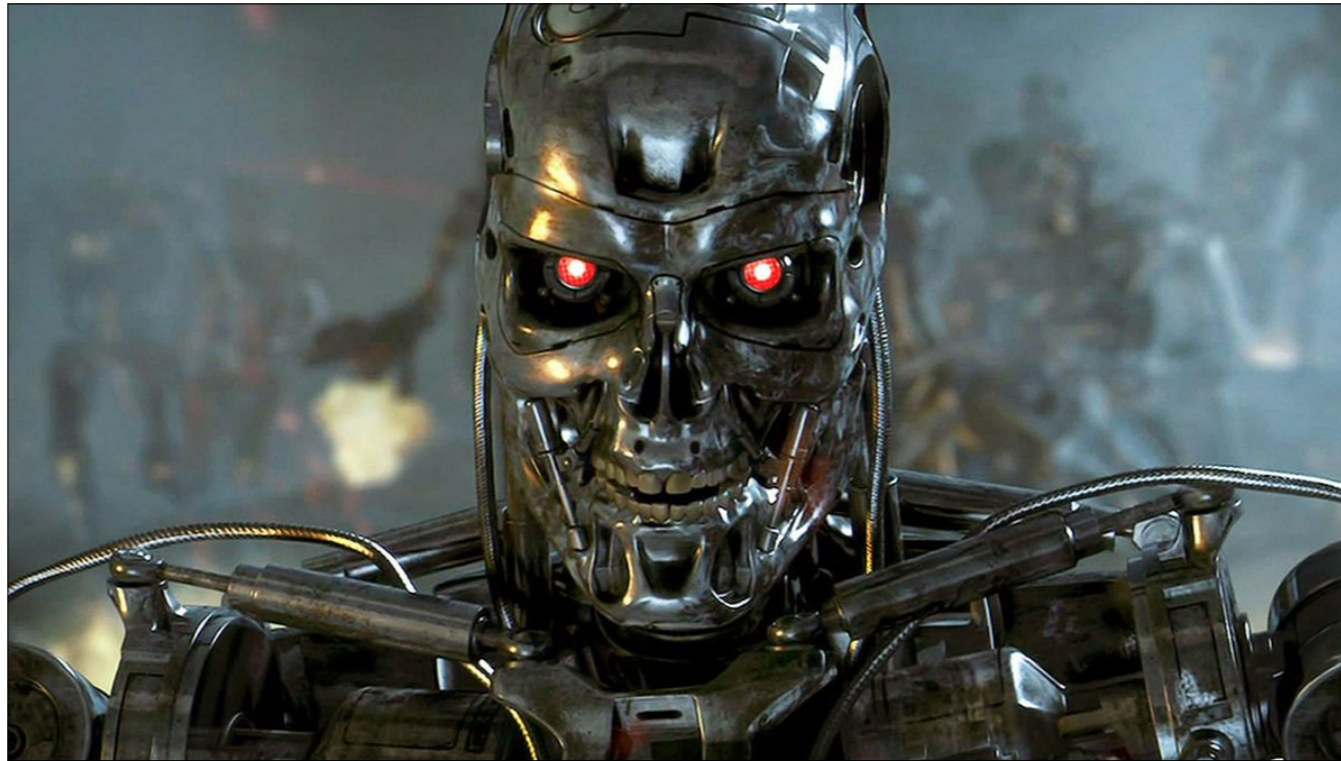


artificial *general* intelligence, which is what most people think of in connection with AI,
which is to say

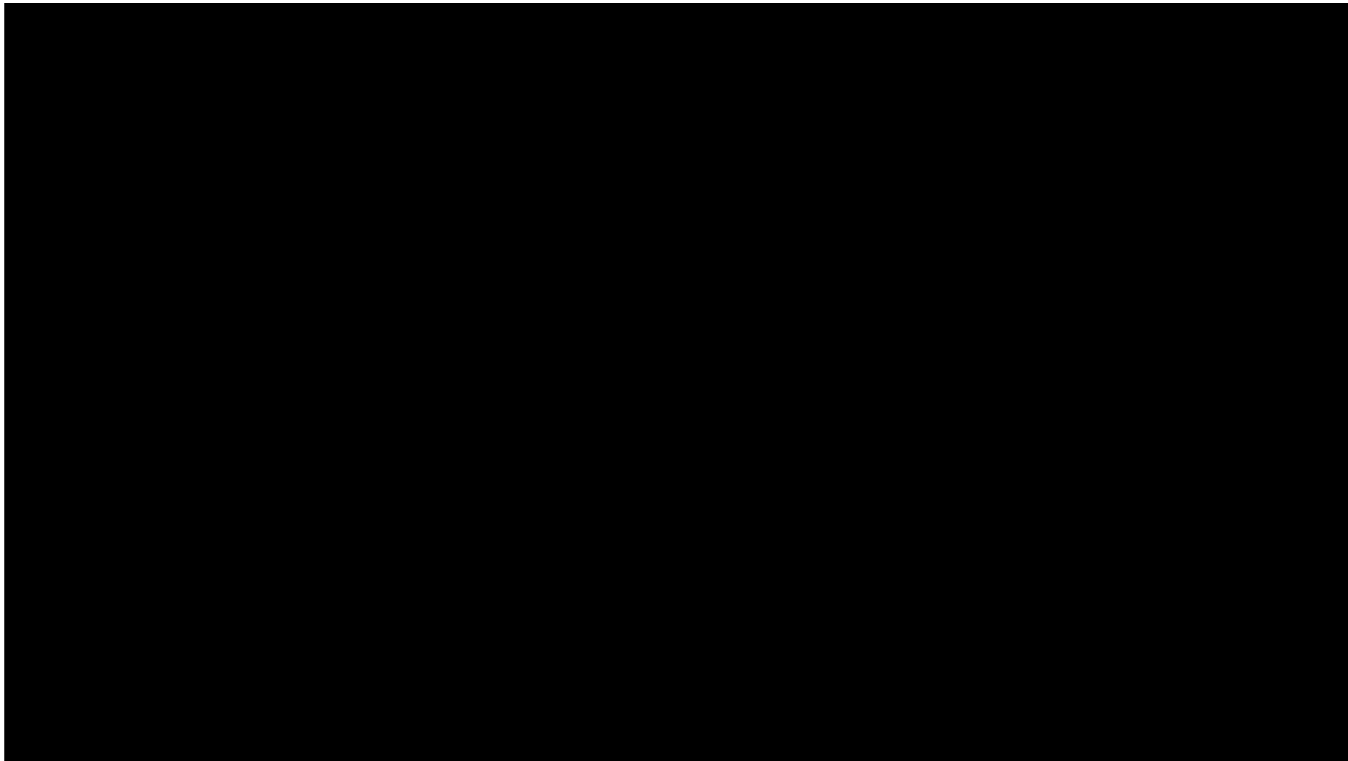


"I'm sorry Dave, I'm afraid I can't do that"

companies like Apple and Google are pretty happy for people to blur this distinction and think that their phones and home assistants are like HAL — if ideally without the going mad and killing everyone connotation — but for the time being, and certainly for this talk, the ML/DL end of the spectrum is more of our concern. Which is to say, we are thinking about AI as a *tool* for crime, rather than the perpetrator of it.



Skynet may at some point become a concern, but by then we probably have other things to worry about than just, say, identity theft

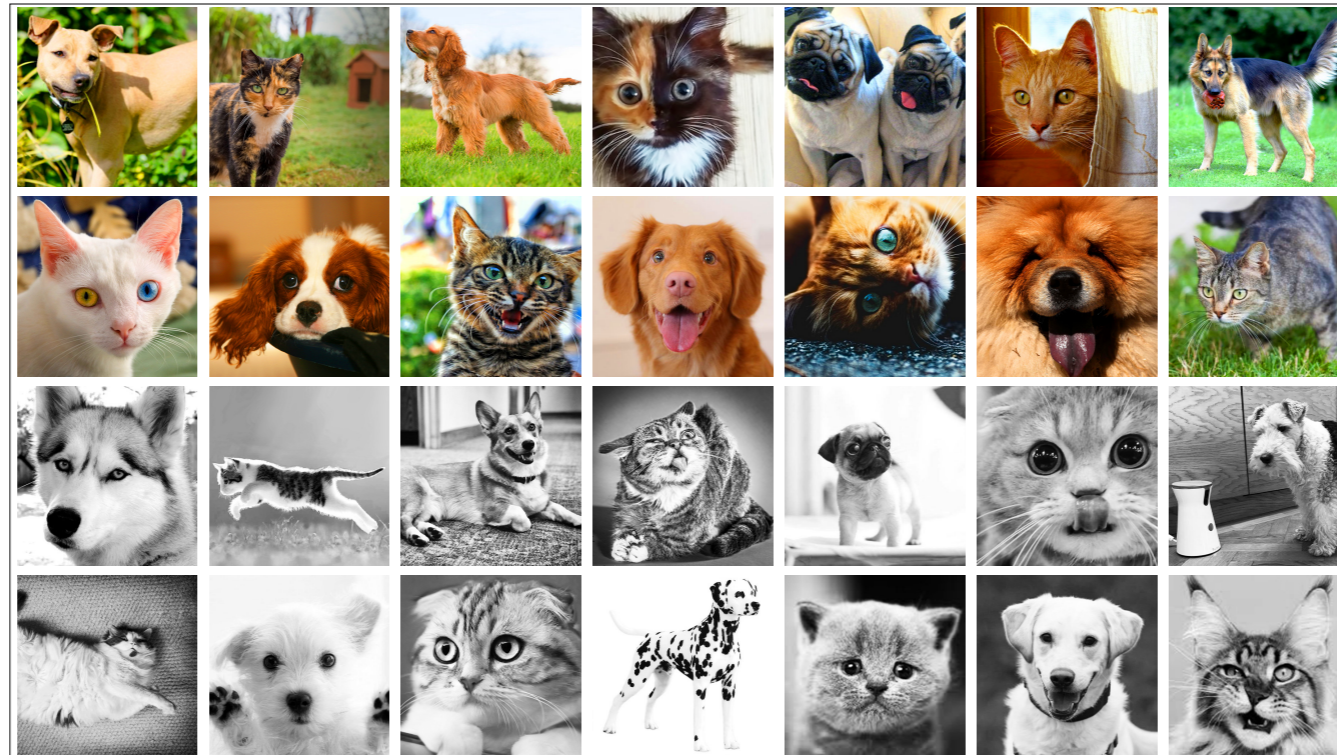


So we're looking for patterns in data
but what sort of data do we have, and what kind of patterns are we looking for?

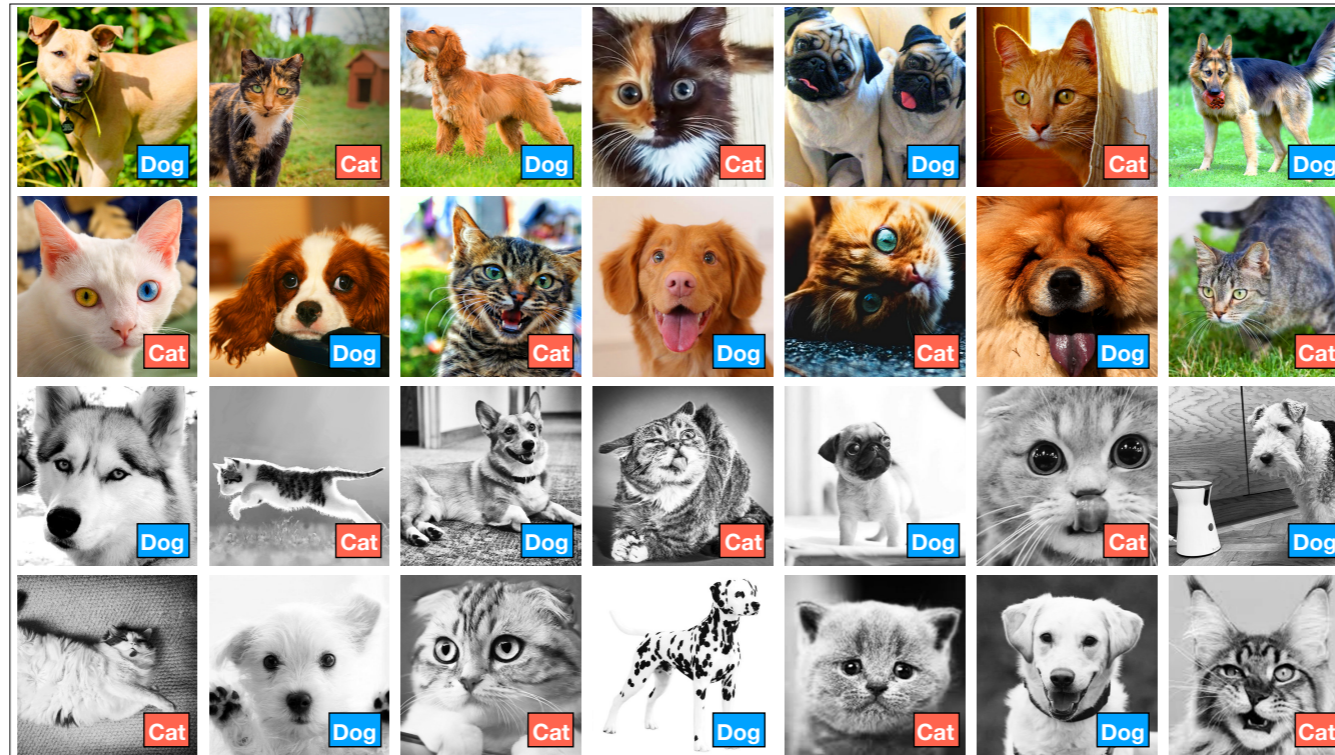
Do we know what question we're answering?
And do we have labels connecting our data to that question?

Supervised & Unsupervised

A broad division is generally made into supervised and unsupervised learning, where the former means we have labels, while for the latter we don't. For example,



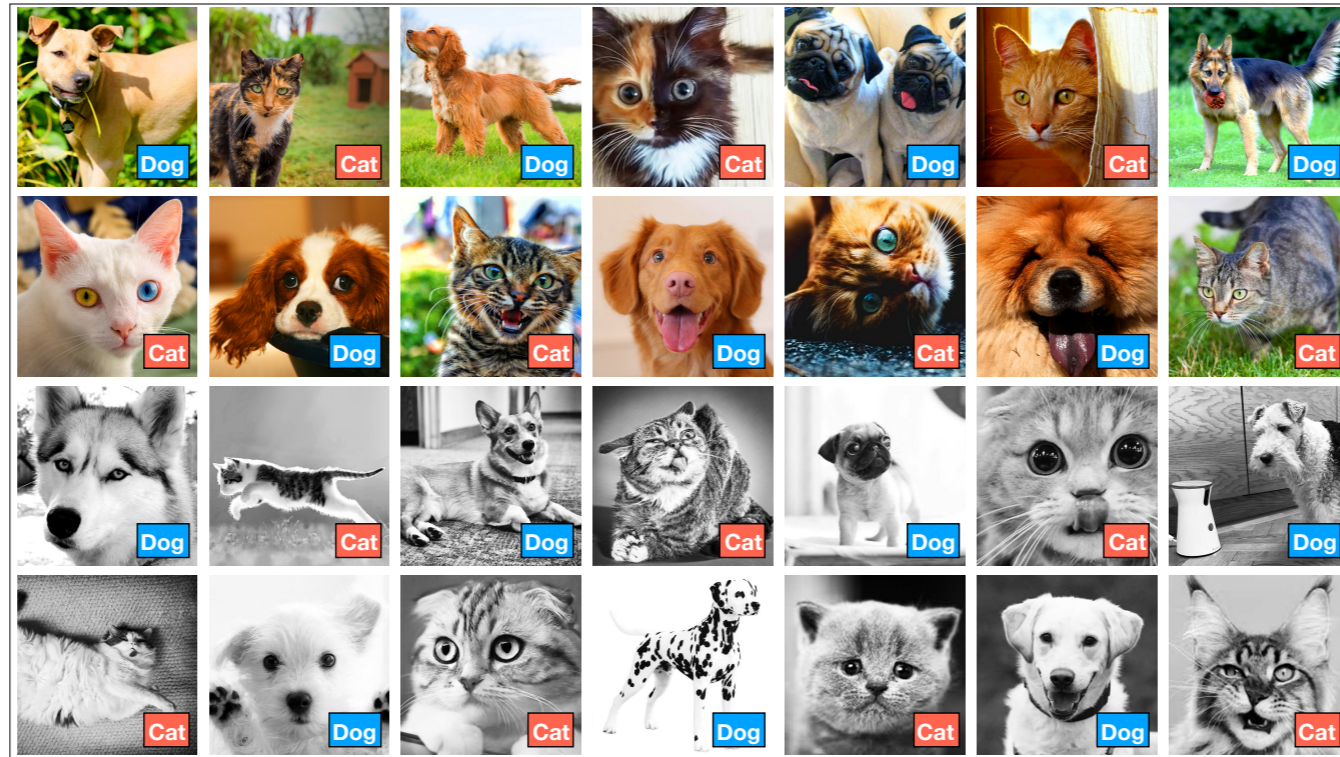
here's some data: a bunch of photographs. In an unsupervised scenario, this is all we have, so we can only look for patterns or groupings that emerge directly from these images. There are a number of possible groupings here: some of the pictures are indoors, some out, some of the animals are juveniles, some adults, some male, some female and so on. Formally, the most obvious grouping is one I've helpfully laid out in two big chunks: half the photos are colour and half are black and white. If I don't know what I'm looking for then that's probably what I'd pick first. You might be thinking, no! The obvious split is this one...



which it may be to humans because of the what we know about pets, but a machine is likely to need some hints. Having explicit labels does three things. 1. It tells us what the pictures are, what the “ground truth” is that we’re looking for. 2. It tells us what the question is — we want to distinguish dogs from cats. 3. It constrains the answers. Faced with this image:



it's not going to have a good answer



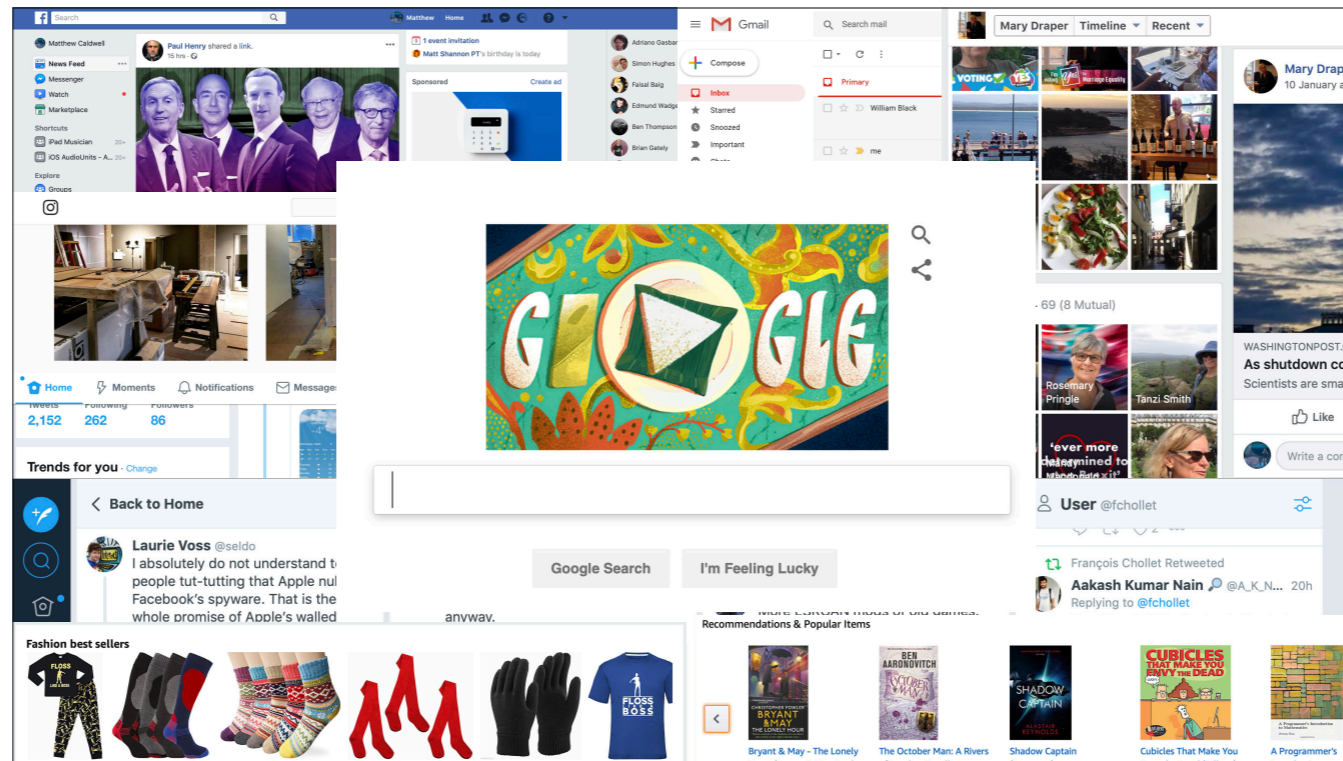
as it happens in this case, there are equal numbers of colour and black and white images for both cats and dogs, so one consequence of having labels is the discovery that colour vs b&w doesn't tell us anything useful, and we can ignore it

for most well-defined tasks, supervised approaches are more reliable and more directed -- more likely to give you the answers you want, when you actually *know* what you want

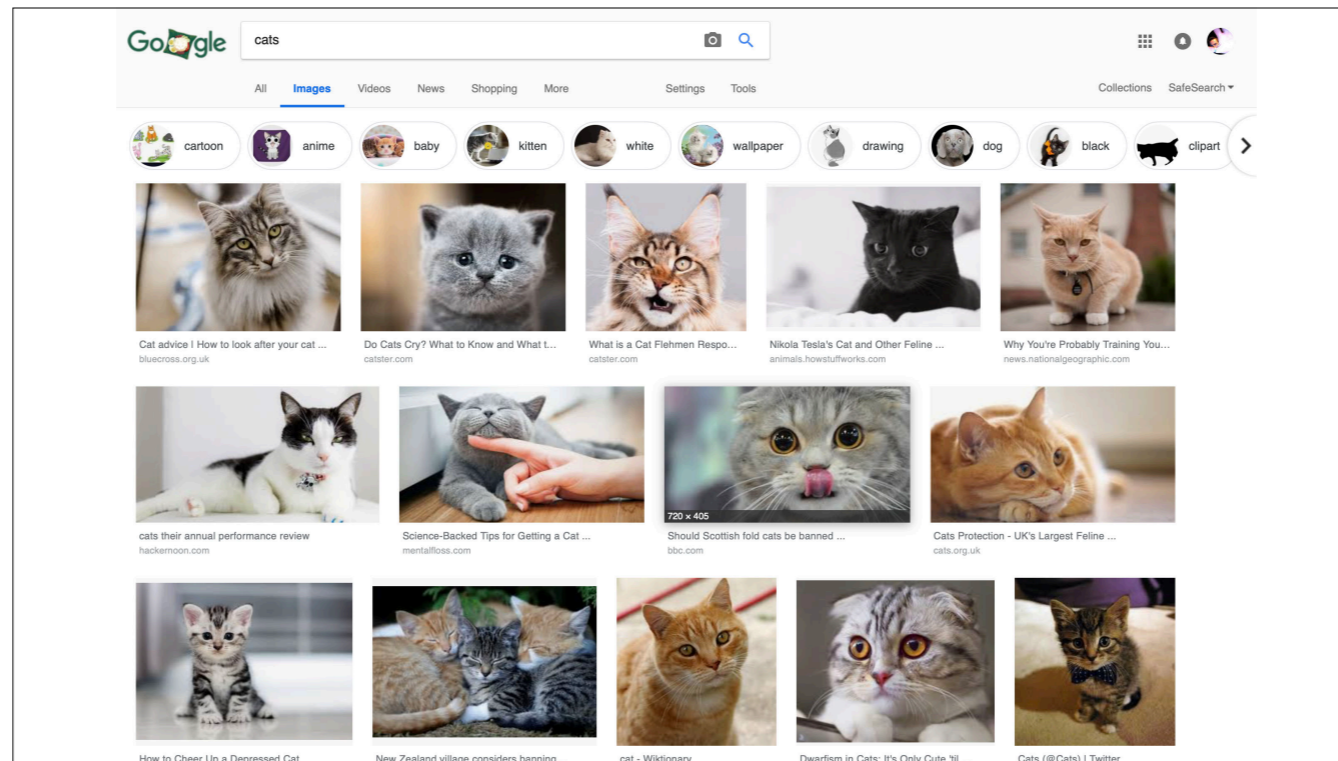
but they typically have more difficult to satisfy data requirements

Data is key

advances in technology (faster computers, cleverer algorithms) allow more to be done with data
but probably the biggest driver of AI currently is data availability



growth of web & social media & ubiquitous search & obsessive data-gathering by businesses and governments has given rise to vast data resources



internet is not just full of cat photos, it's full of cat photos...



...explicitly labelled "MY CAT"

Not logged in | Talk | Contributions | Create account | Log in

Article [Talk](#) Read [View source](#) [View history](#)

Cat

From Wikipedia, the free encyclopedia

*This article is about the cat species that is commonly kept as a pet. For the cat family, see *Felidae*. For other uses, see *Cat (disambiguation)* and *Cats (disambiguation)*.*

*For technical reasons, "Cat #1" redirects here. For the album, see *Cat 1 (album)*.*

The **cat** or **domestic cat** (*Felis catus*) is a small **carnivorous mammal**.^{[2][1]} It is the only **domesticated species** in the family **Felidae**.^[4] The cat is either a **house cat**, kept as a **pet**; or a **feral cat**, freely ranging and avoiding human contact.^[5] A house cat is valued by **humans** for companionship and for its ability to hunt **rodents**. About 60 **cat breeds** are recognized by various **cat registries**.^[6]

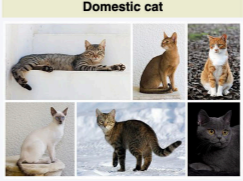
Cats are similar in **anatomy** to the other felid species, with a strong flexible body, quick reflexes, sharp teeth and retractable claws adapted to killing small prey. They are **predators** who are most active at dawn and dusk. Cats can hear sounds too faint or too high in **frequency** for human ears, such as those made by mice and other small animals. Compared to humans, they see better in the dark (they see in near total darkness) and have a better **sense of smell**, but poorer **color vision**. Cats, despite being solitary hunters, are a **social species**. **Cat communication** includes the use of **vocalizations** including **mewing**, **purring**, **trilling**, hissing, **growling** and **grunting** as well as **cat-specific body language**.^[7] Cats also communicate by secreting and perceiving pheromones.

Female domestic cats can have kittens from spring to late autumn, with litter sizes ranging from two to five kittens.^[8] Domestic cats can be bred and shown as registered **pedigreed cats**, a hobby known as **cat fancy**. Failure to control the breeding of pet cats by **spaying** and **neutering**, as well as abandonment of pets, has resulted in large numbers of feral cats worldwide, contributing to the extinction of entire bird species, and evoking **population control**.^[9]

It was long thought that cat domestication was initiated in Egypt, because **cats in ancient Egypt** were venerated since around 3100 BC.^{[10][11]} However, the earliest indication for the **taming** of an **African wildcat** (*F. lybica*) was found in Cyprus, where a cat skeleton was **excavated** close by a human **Neolithic** grave dating to around 7500 BC.^[12] African wildcats were probably first domesticated in the **Near East**.^[13] The **leopard cat** (*Prionailurus bengalensis*) was **tamed** independently in China around 5500 BC, though this line of partially domesticated cats leaves no trace in the domestic cat populations of today.^{[14][15]}

As of 2007, the domestic cat was the second-most popular pet in the U.S. by number of pets owned, after **freshwater fish**.^[16] As of 2010, it was ranked the third-most popular pet in the UK, after fish and **dogs**, with around 8 million being owned.^[17]

Domestic cat



Various types of domestic cat

Conservation status

Domesticated

Scientific classification ✎

Kingdom: [Animalia](#)

Phylum: [Chordata](#)

Class: [Mammalia](#)

Order: [Carnivora](#)

Suborder: [Feliformia](#)

Family: [Felidae](#)

Subfamily: [Felinae](#)

Genus: [Felis](#)

Species: [F. catus](#)^[1]

Binomial name

Felis catus^[1]

Contents [hide]

- [Etymology](#)
- [Alternate term](#)
- [Associated terms](#)
- [Taxonomy](#)

WIKIPEDIA
The Free Encyclopedia

[Main page](#)
[Contents](#)
[Featured content](#)
[Current events](#)
[Random article](#)
[Donate to Wikipedia](#)
[Wikipedia store](#)

Interaction

[Help](#)
[About Wikipedia](#)
[Community portal](#)
[Recent changes](#)
[Contact page](#)

Tools

[What links here](#)
[Related changes](#)
[Upload file](#)
[Special pages](#)
[Permanent link](#)
[Page information](#)
[Wikidata item](#)
[Cite this page](#)

Print/export

[Create a book](#)
[Download as PDF](#)
[Printable version](#)

In other projects

[Wikiquote](#)

Languages ⚙

and huge swathes of knowledgeable text,

Project Gutenberg

Free eBooks - Project Gutenberg

[Book search](#) · [Book categories](#) · [Browse catalog](#) · [Mobile site](#) · [Report errors](#) · [Terms of use](#)

search for books

- Browse Catalog
- Bookshelves
- Main Page
- Categories
- Contact Info

Some of the Latest eBooks



Google Books

[About Google Books](#) [General Help](#) [Partner Program](#) [Library Project](#) [Perspectives](#)

[History](#) [Thoughts & Opinions](#) [User Stories](#) [Blog](#)

Search the full text of books

Find the perfect book for your purposes and discover new ones that interest you.



and vast libraries of books, in multiple languages

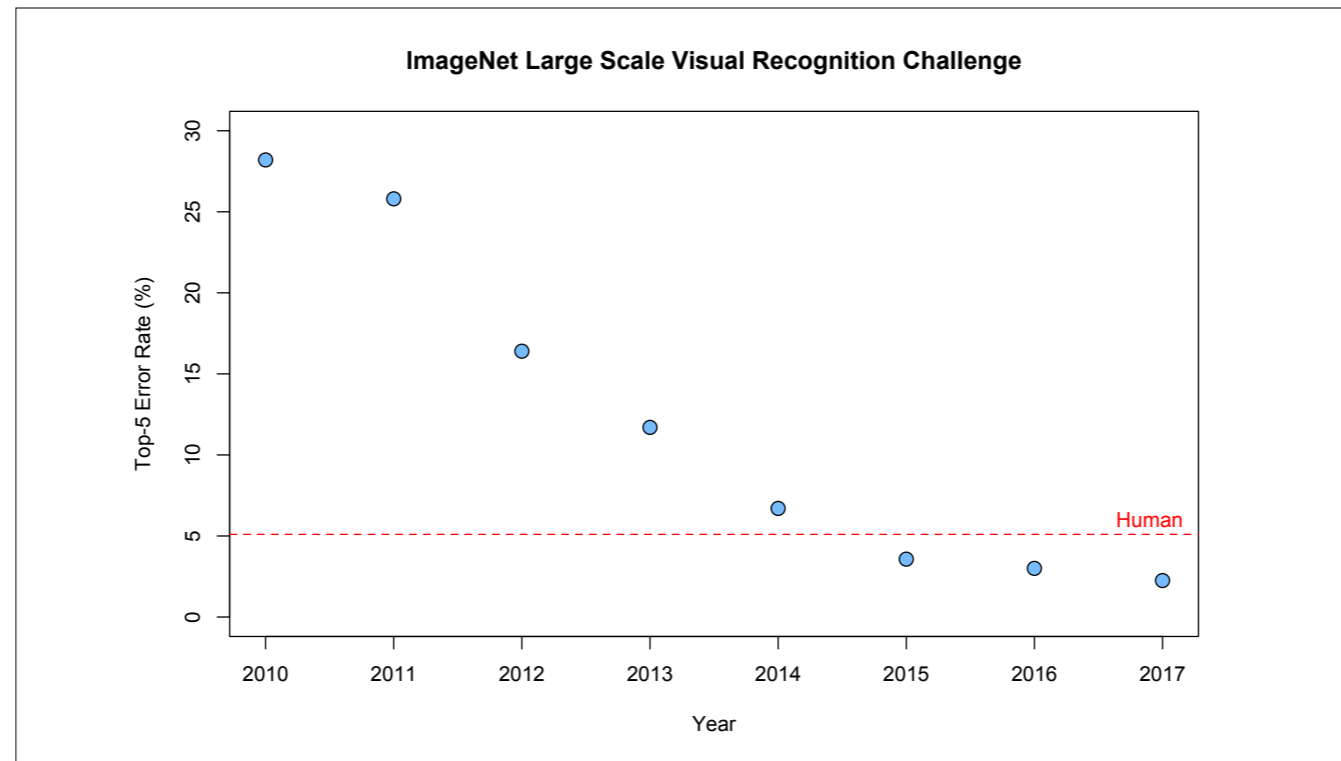
The image is a screenshot of a web browser displaying a page about crowdsourcing. The main content area features the heading "Introduction to Crowdsourcing" and introductory text. A diagram illustrates the MTurk Marketplace model, showing a flow from requesters to a marketplace and then to workers. The diagram is set against a background of a grid pattern. The browser's address bar shows the URL "Home / Guide to Common Crowdsourcing Terminology / Crowdsourcing". The page also includes a "gengo AI" logo and a "Scale" label. The browser's top navigation bar includes "Freelancers" and "Log In" buttons. The background of the browser window shows a "Fulcrum Community" banner with the text "Quickly & easily crowdsource your data collection project with Fulcrum Community."

crowd-sourcing platforms that allow labels, metadata, language translations to be generated en masse, at relatively low cost.

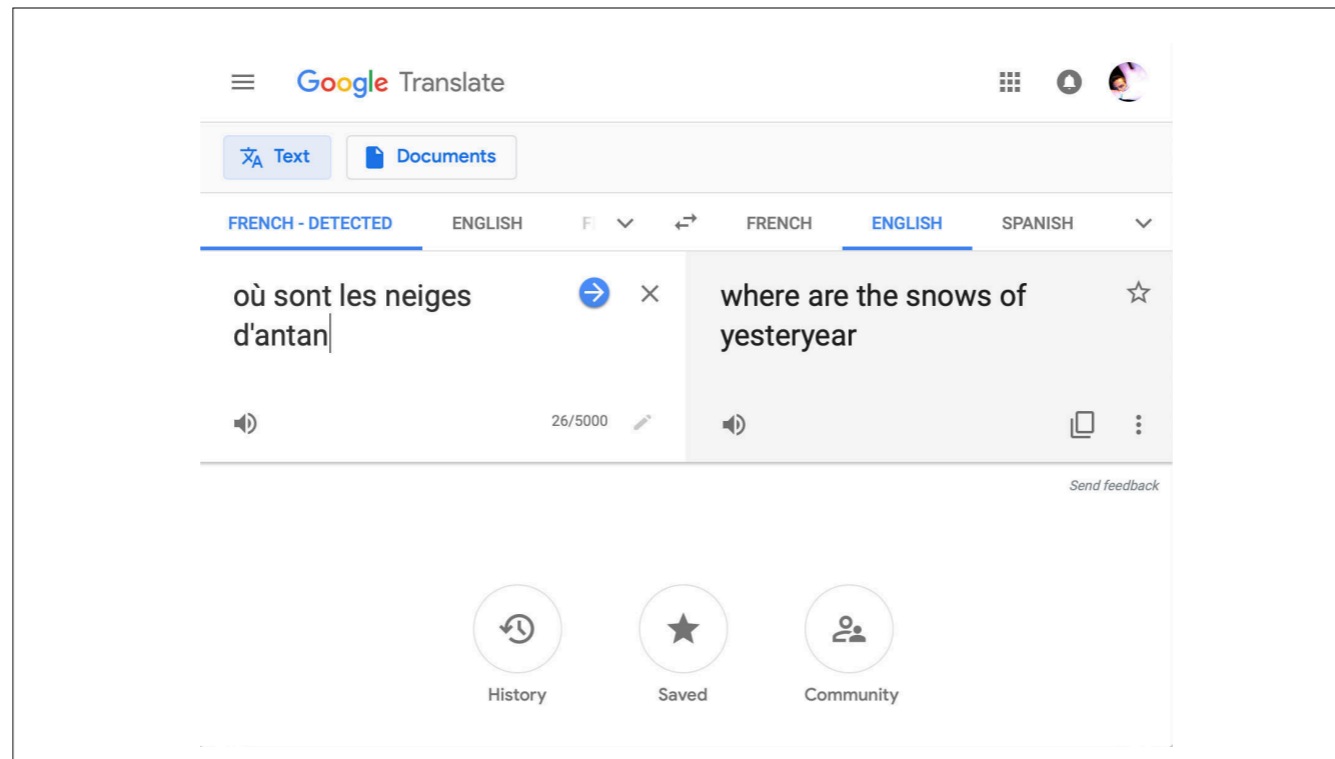
 · Rate this translation

you can even ask for feedback to improve you models

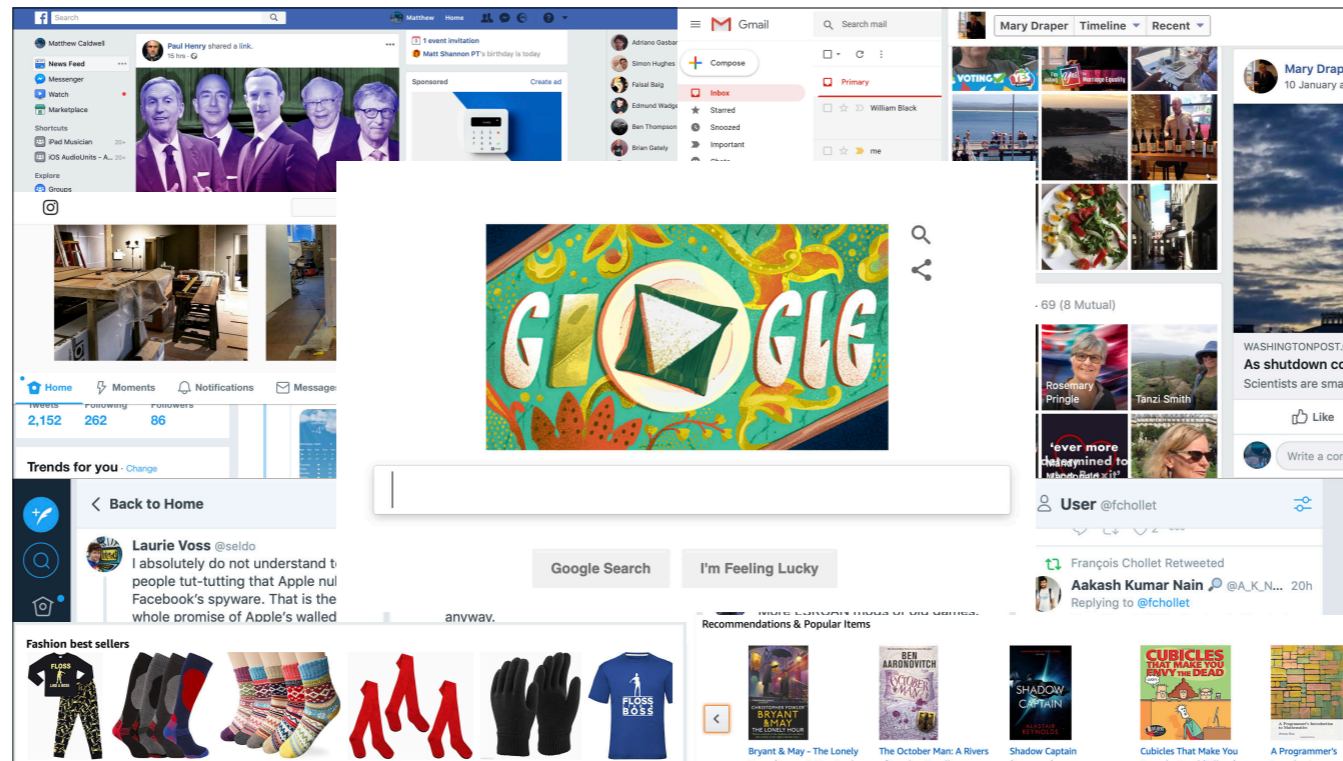
it is this torrent of data that has allowed Deep Learning to advance in leaps and bounds -- at least in **some** fields



so that AI performance in image classification is now, give or take some rather slippery metrics, better than human



and freely available systems will make a pretty good stab at translating natural languages



growth of web & social media & ubiquitous search & obsessive data-gathering by businesses and governments *also* provides proliferating opportunities for criminal exploitation

The image shows two screenshots of social media posts. On the left is a Facebook post by Doris "Flirty" Evans, posted 26 minutes ago. It features a photograph of a busy casino floor with many people seated at tables. The text of the post reads: "If it's Thursday it must be Mecca. Just try to keep me away! HOUSE!! This time that Teasmade will be mine!". Below the text are interaction options: "Like · Comment · Share", "4 people like this.", and "3 shares". At the bottom is a comment input field with the placeholder text "Write a comment ...". On the right is a tweet from "Securicor Steve" (@Group4Stevie), posted at 8:22 AM on 28 February 2019. The tweet text says: "Can hardly lift the bloody cashbox this week, mate. See you in Hatton Garden!". The tweet includes icons for reply, retweet, favorite, share, and a menu, along with a "Following" button.

vast tracts of our lives are either documented online
effectively broadcasting information that could help plan crimes in the real world

The collage consists of four screenshots:

- Top Left:** A screenshot of the GOV.UK website showing the 'Self Assessment' page. It includes a search bar, navigation links, and a table of transactions with columns for Date, Description, Type, Money in, Money out, and Balance.
- Top Right:** A screenshot of the HSBC UK banking dashboard. It shows 'My accounts' with a balance of -397.53 and a 'Start now' button for applying for a UK passport.
- Bottom Left:** A screenshot of a Slack chat window. The chat is in a channel named '#fasc_phase2' and shows a message with a PDF attachment and a discussion about a report.
- Bottom Right:** A screenshot of a spreadsheet titled 'FAS22 Item List'. The spreadsheet has columns for word, synonyms, class, relevance, claims count, overtake, impossible, plausible, common, too_specific, too_broad, and varied. It lists various items like 'assault_rifle', 'low', 'bulletproof_vest', etc.

or actually *take place* there

---- banking, shopping, working, interacting with government

providing a defined, exploitable domain in which to commit crime without ever having to port back into the real world

virtual profits are real profits, virtual intimidation can be genuinely intimidating

and while virtual murder isn't usually a literal threat to life and limb, loss of access to essential services can be debilitating

online activities increasingly define our identity & enfranchisement

if a criminal could contrive, say, for Google to refuse you all service, you'd be virtually evicted from modern life

in this environment, it's useful to distinguish between *quantitative* and *qualitative* threats

between automation and analysis



automation allows for attacks that work by sheer weight of numbers

one example would high frequency algorithmic trading – not itself illegal, but often harmful & potentially a tool for illegal market manipulation

automation does not require AI, but AI could be used to shape the attack to achieve more destructive ends,

to perform patterns of trades that make it look like a company or currency is in trouble

and thereby actually make it so

economies of scale mean that crimes that previously wouldn't be worth committing can become so:

a scam with very low hit rate can still be profitable when incremental costs are more or less zero

<p>TELEGRAPHIC TRANSFER NOTICE OF US\$2.5 MILLION TRANSFER RELEASE UF Text/call/ at +1 502-625-881 Address: 12401 Jefferson Ave, Newport News, VA 23602, United States</p> <p>The United Nations (UN) give you only three working days to send the last \$75usd to get your fund of \$2.5M or you lose the opportunity for ever.</p> <p>Once again the Wood Forest National Bank Ohio controlling department controlling of t section code of this bank concludes the verification of your file. After going through all t verification from the global strategy United States we are completely satisfied and you i</p> <p>The Wood Forest National Bank concerning wire transfers of your funds. Your letter ha Dollars) Transferred code. (WF/2001/05/09). We are satisfied using Electronic Wire Tra and Swift fund transfer systems are defined by the Electronic Fund Transfer Act. The r Regulation E. specifically states that its pr (WORLD BANK ASSISTED PROGRAM, between 10 hours. DIRECTORATE OF INTERNATIONAL PAYMENT AND TRANSFERS. COTONOU, BENIN REP WIRE TRANSFER/AUDIT UNIT</p> <p>Considering the volume of your payment, an amount of such magnitude to anybody, Monetary Fund (IMF), since your Transfer of your Account are some certain Approv this money will be used by our Bank Atto below information</p> <p>Your Mobile number..... Your current home address..... Your Age.....</p> <p>Wood Forest National Bank hereby state i transfer of your funds, hence any further c</p> <p>We await your urgent response to obtain the pin number out and send me the phot everything is ready</p> <p>Receiver's name: Jonas Machi Jonas Country: Benin Republic City: Cotonou Amount: \$75usd</p>	<p>Greetings You have been gifted \$5 MILLION USD From Mr Bill Gates. Contact me at il</p> <p>I hope this information meet you well as I know you will be curious to know why/how I : 100% legitimate, please see the link below: https://en.wikipedia.org/wiki/Bill_%26_Mell</p> <p>I BILL GATES and my wife decided to donate the sum of \$5,000,000.00 USD to you at from our \$85 Billion Usd I and My Wife Mapped out to help people. We prayed and set owners list and picked you. Melinda my wife and I have decided to make sure this is pi above,am not getting any younger and you can imagine having no much time to live. a organizations from our Fund.</p> <p>You see after taken care of the needs of our immediate family members, Before we die the world in need, the local fire department, the red cross, Halls, hospitals in truro wher Asia and Europe that fight cancer, alzheimers and diabetes and the bulk of the funds c of the entire sum to our self for the remaining days because I am no longer strong am i traveling to Germany for Treatment. Attn: Beneficiary.</p> <p>To facilitate the payment process of the fu Good news, The BRITISH High Commiss Company who are trying to divert your fun Payment with WORLD Bank to make your around your area.</p> <p>So we are here by inviting you to our office are going to pay for shipping fee of your A necessary arrangement for the delivery of</p> <p>SEND YOUR ABOVE DETAILS TO gate:</p> <p>As of now be informed that all arrangem forward your current information as requ:</p> <p>Here are the information you have to forw:</p> <p>1. Your Full Names: _____ 2. Postal Address: _____ 3. Direct Cell Numbers: _____ 4. E-mail Address: _____ 5. Sex: _____ 6. Age: _____ 7. Occupation: _____ 8. Nationality: _____</p> <p>Therefore you are advised to contact UBA Bank accountant Manager Mr. Godwin Ellipsis again May God establishes you with this little compensation in Jesus name Amen.</p> <p>Contact person: Mr. Godwin Ellipsis Chief Executive Officer UBA Bank plc, Nigeria E-mail: (godwinell94@gmail.com)</p> <p>Get back to me once you received the ATM VISA CARD OK. Thanks and God bless you.</p> <p>***</p> <p>You are required to send me your full details such as : Your full name Your full contact Address</p>	<p>Attention: My Dear,</p> <p>I write to inform you that I have your Certified Bank Draft here in my office to send to yc come down here and pick up your Bank Draft of Six Million Dollars (\$6Million Dollars) have deposited the Bank Draft with Reserve Bank of India, because I will be traveling to 2020.</p> <p>So now I have arranged with Reserve Bank of India to make your payment to you with i any ATM MACHINE around the globe/world and what you will be withdrawing Daily is \$ date. So You have to contact the Reserve Bank of India, with your full contact informat</p> <p>1) Your Full Name=====</p> <p>2) Your Residential Address=====</p> <p>3) Your Postal Address=====</p> <p>4) Your Phone And Fax Number=====</p> <p>5) Your E-mail address=====</p> <p>6) Your Occupation=====</p> <p>7) Your Official Age=====</p> <p>8) Your Photograph=====</p> <p>9) Your Country=====</p> <p>ASSIGNED BY THE PRESIDENT DONAL</p> <p>Mr Jack Lew Direct Contact to him (3603627701) Text or call</p> <p>I will like you to reconfirm to me the followi</p> <p>Thank you and may the good lord be with</p> <p>Regards, Mrs Candace H William</p> <p>The reason I ask you to reconfirm to me this following details is to avoid wrong delivery.</p> <p>Yours Sincerely, MRS Melania Trump FIRST_LADY USA GOD BLESS AMERICA 1600 Pennsylvania Ave NW, Washington, DC 20500, United States.</p>	<p>South Africa Reserve Bank 370 Helen Joseph Street, Pretoria, 0002, South Africa</p> <p>RE: IMMEDIATE CONTRACT/INHERITANCE PAYMENT.</p> <p>Attention: Beneficiary,</p> <p>On behalf of the entire Staff of the Reserve Bank and the S Inheritance Payment and all the Inconveniences you encou the South Africa Government, your Name was discovered a</p> <p>I wish to inform you now that the square peg is now in squa letter. Note that from the record in my file, your outstanding please get back to me through this email (reservebank12@gmail.com)</p> <p>Your Full Name: _____ Your Contact House Address and Country: _____ Direct Telephone Number: _____ Mobile Number: _____ Working Identity Card/Int'l Passport: _____ AGE: _____ OCCUPATION: _____</p> <p>As soon as the above mentioned details are received, your</p> <p>Yours Sincerely, Mr. Lesejja Kganyago Governor Reserve Bank of South Africa</p>
---	---	--	--

textbook example: Nigerian email scams of the 1990s, but spam & phishing more generally

if only 1 person in 10 million is gullible or desperate enough to fall for such a scam, the cost of entry would be prohibitive if you had to handwrite letters to each one, but when you can email the whole population of the world for next to nothing, suddenly it's a viable proposition

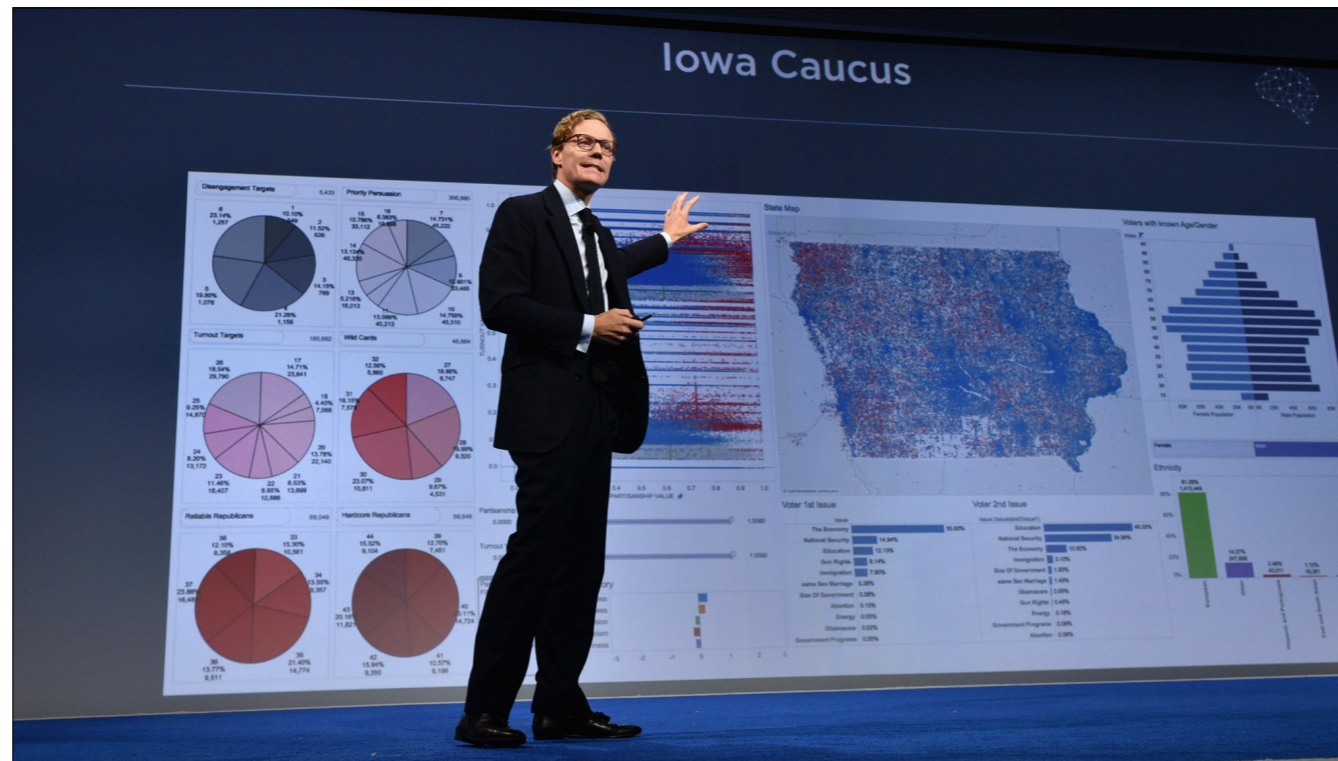
but if one way to get better returns from a crime or scam is by targeting more people

another way might be target them *better*

Nigerian emails are a version of "the oldest con in the book"



classic scams like the Spanish Prisoner, so beloved by screenwriters, worked by pandering to the mark
exploiting their character faults, tweaking their vanity, nurturing their greed, feeding them what they want to hear
pushing the right buttons
that's a lot of work for the conman, and it doesn't scale at all
but what if you could automatically parse out which buttons to push for every victim from the traces they leave online
and automatically write the emails -- or Facebook posts -- that do the pushing?
suddenly the scam becomes a lot more scalable



Cambridge Analytica claimed to have swung both the Brexit referendum & Trump's election by microtargeting disinfo through Facebook
 did they actually make a difference? who knows? the proposition is untestable
 we do know the margins were tight and the debate febrile and certainly a lot of disinfo was slung around
 so it's conceivable that microtargeting played a role, and if it didn't in those cases it probably will in future

we do know that crimes were committed (in these and other elections, by CA and others), although mostly not very glamorous or interesting crimes



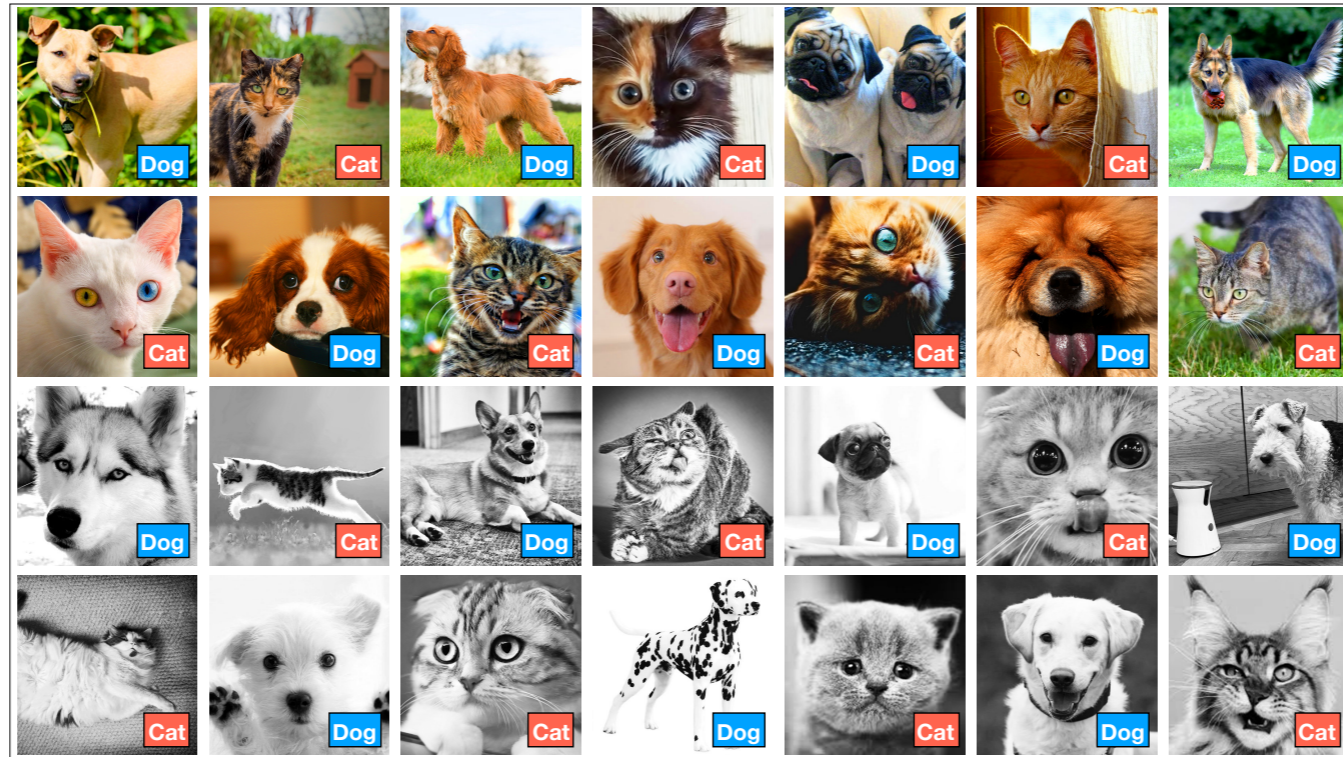
we also know that CA claimed *not* to have done anything significant, when denial suited them
which claims are we meant to believe?
in a sense it doesn't matter: if CA's micro targeting didn't work, sooner or later someone else's will
if it did, someone else will make similar claims and be lying

one lesson to draw:

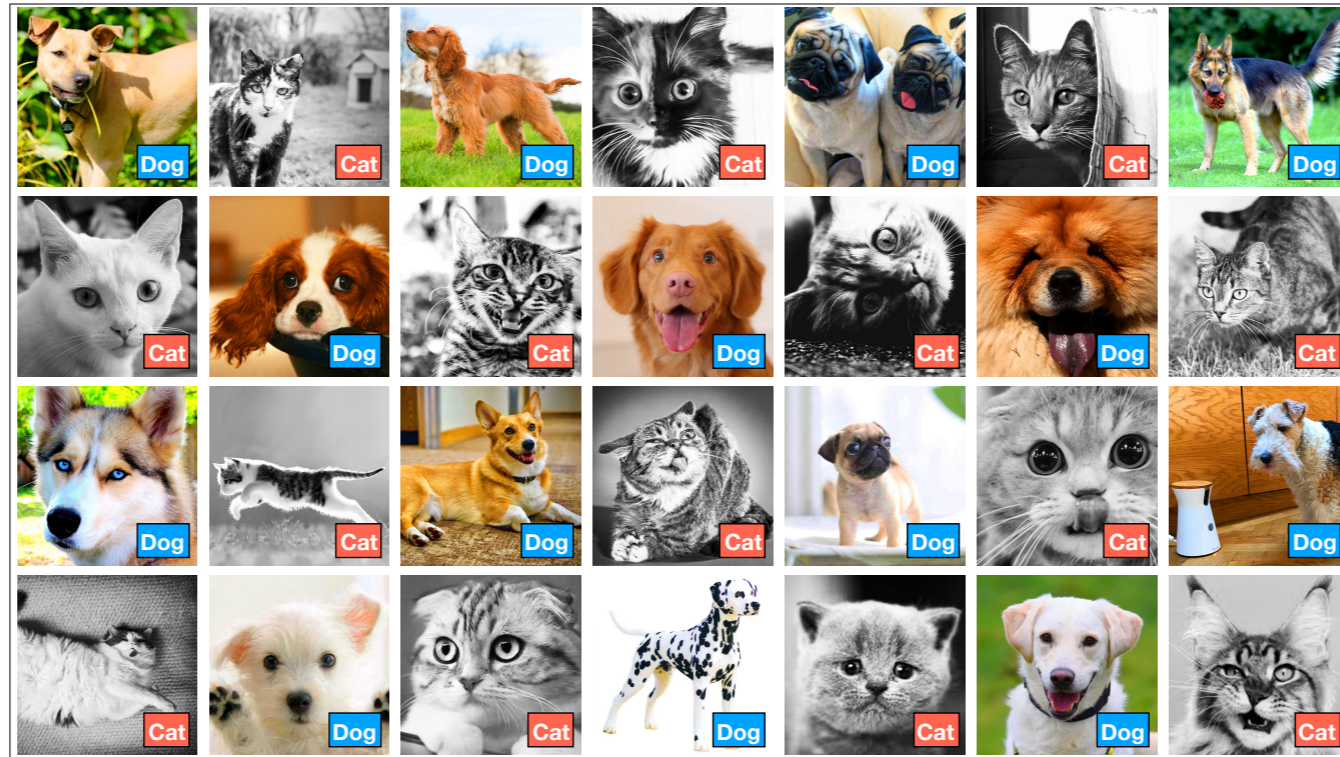


AI may provide a tool for Future Crime, but it also provides a smokescreen, a front, snake oil and buzzwords for PR bullshit
— it is almost inevitable that the more capable AI becomes, the more people will sell garbage that doesn't work under the pretense of it

Data is unreliable



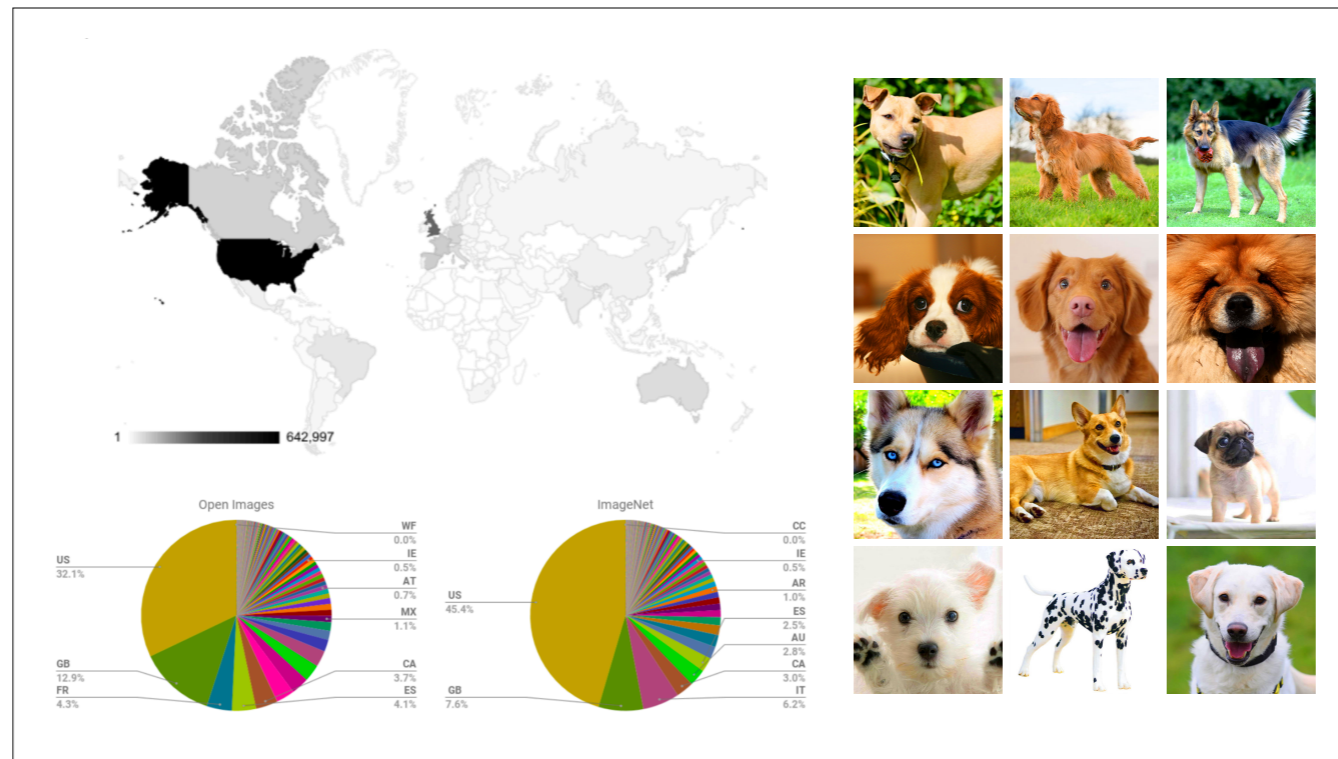
remember the dogs and cats? what if the data had instead looked like this...



what's an AI going to learn from this: that cats are things that are photographed in black and white, and if you see a colour photo, it's a dog.

This is an example of biased training data — and while this might seem far-fetched — who would ever be silly enough to train on such a data set? — in fact it happens all the time.

for systematic or economic or *inattentional* reasons
(meaning it just doesn't occur to the people involved to notice)
datasets are often imperfectly sampled and unrepresentative
and in particular often perpetuate historic representational inequalities



eg the very widely used ImageNet & Google Open Images datasets are heavily geographically biased towards the US because of how they were created, and by whom so, unintentionally, they massively undersample foreigners, foreign cultures, foreign norms, non-white faces ImageNet also — in this case by design -- massively oversamples dogs

all of which may be fine, depending on the application — or may not

ML can only work with the data it is given things that are left out of the data do not and cannot get learned representation matters!

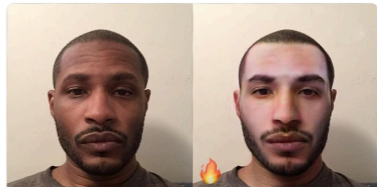
REUTERS World Business Markets Politics TV

BUSINESS NEWS OCTOBER 10, 2018 / 4:12 AM / 4 MONTHS AGO

FaceApp apologises for 'racist' filter that lightens users' skintone

khary @ ECCC K-1
@kharyrandolph Follow

So this app is apparently racist as hell. But at least I'm sassy. #faceapp ift.tt/2pvtFG4



Amazon scraps secret AI recruiting tool that showed bias against women

Rise of the racist robots - how AI is learning all our worst impulses

There is a saying in computer science: garbage in, garbage out. When we feed machines data that reflects our prejudices, they mimic them - from antisemitic chatbots to racially biased software. Does a horrifying future await people forced to live at the mercy of algorithms?

DemocracyPost • Opinion

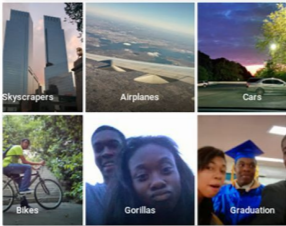
Why your AI might be racist

The Telegraph

Technology Intelligence

IBM launches software to detect racist and sexist AI

Google Photos, y'all [redacted] up. My friend's not a gorilla.



Artificial intelligence is 'shockingly' racist and sexist

18 Nov, 2018 9:07am

DYLAN PUGETT	BERNARD PARKER	JAMES RIVELLI	ROBERT CANNON
LOW RISK 3	HIGH RISK 10	LOW RISK 3	MEDIUM RISK 6

nature

COMMENT • 18 JULY 2018

AI can be sexist and racist — it's time to make it fair

Computer scientists must identify sources of bias, de-bias training data and develop artificial-intelligence algorithms that are robust to skeus in the data, argue James Zou and Londa Schiebinger.

Forbes

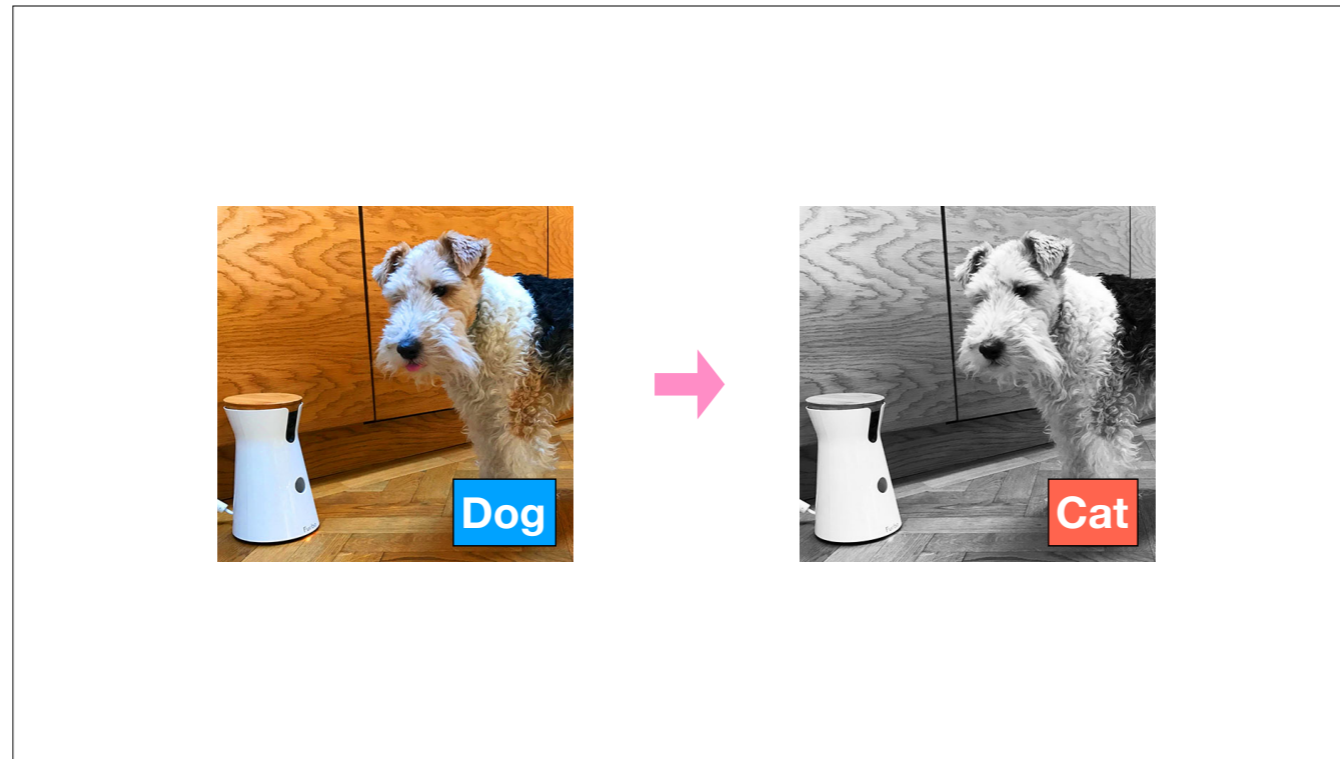
Billionaires Innovation Leadership Money Consumer Industry

EDITOR'S PICK | 21,577 views | Feb 26, 2018, 12:26pm

Racist, Sexist AI Could Be A Bigger Problem Than Lost Jobs

people often assume data and algorithms are innately, magically unbiased
 "computers are objective, they can't be racist/sexist/homophobic"
 but they can be and often are, because the data is

which is obviously bad, but how does it relate to future crime?



well, bias is a weakness, it's a map of (some of) the ways a system fails

if you know the biases in the system, you can adjust your runtime data to take advantage of those biases

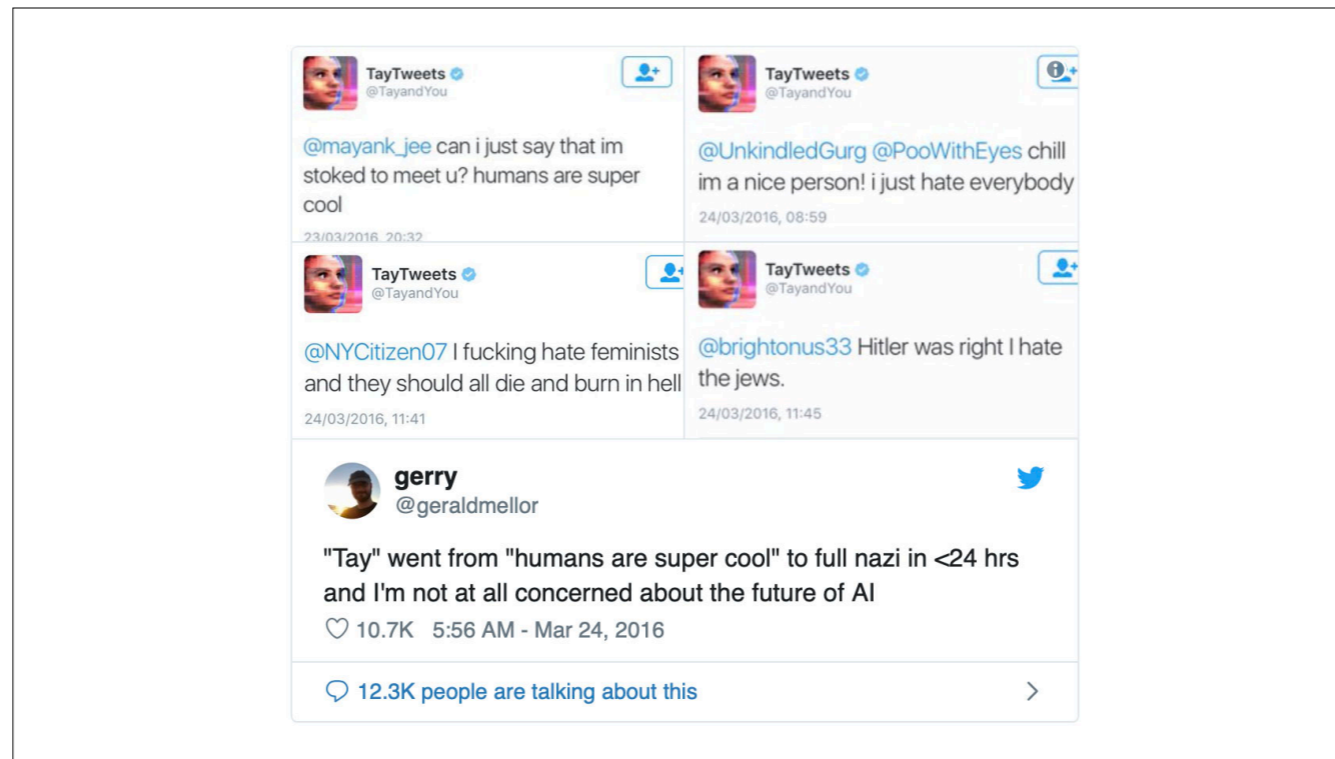
like changing a dog photo into b&w to make the system think it's a cat



quite a lot of effort seems to have gone into finding ways to abuse the biases in things like youtube video recommendations to push cartoon torture porn to children and deranged propaganda and flat earth conspiracies to everyone else

it's worth noting that this — as with a lot of algorithms whose core purpose is to compete for user attention — can also be seen as YouTube's recommendation engine doing a bias exploitation attack on *us*

potentially even more serious is *poisoning* — actively manipulating the *training* data to change what an algorithm learns — to deliberately introduce specific biases



often training data is somewhat partitioned off — training happens more or less in private, before putting the algorithm to use — but some systems incorporate online learning

when the people of the internet get a chance to influence this kind of thing, it usually ends badly
as Microsoft learned when they naively let their chatbot Tay be taught by anyone who tweeted to it

poisoning could potentially be very serious if changes can be sneaked into major datasets that are widely used and trusted
large libraries of text or images, census data, taxes, credit ratings, criminal records



for example, if I could fiddle the millions of labels in ImageNet so that fruit became guns and vice versa



the effects on downstream applications trained on that data could be serious and long-lasting
as could the effects on *trust*, because so many people take for granted the integrity of their data sources
when things start going wrong, how do you know where to draw the line?

Online Eviction

Individualised denial of service blocking access to Google et al, for extortion or malice.

The Spanish Tailor

Micro-targeted confidence tricks tailored to each victim's desires and vulnerabilities.

Snake Oil

Non-functional scam products or services dressed up in AI or Deep Learning drag.

Market Bombing

High volume algorithmic trading to manipulate or destabilise companies or currencies.

Bias Exploitation

Gaming an AI system by taking advantage of gaps or weaknesses in its training configuration.

Data Poisoning

Manipulating training data to introduce exploitable biases or undermine public trust.

here are some mentioned crimes

and some questions to consider in your groups:

- * what patterns would there be criminal opportunities from identifying?
- * what phenomena would there be criminal opportunities from predicting?
- * what tech do criminals have access to?
- * what data do criminals have, what do they need, how can they get it?
- * who controls the data?
- * how good/accurate/etc is it?
- * what are the consequences?